SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	4/20/99	Final Progress Report 1995-1998

4. TITLE AND SUBTITLE
Prediction of Reduced Ion Mobility Constants from Molecular Structure

5. FUNDING NUMBERS
DAAH04-95-1-0597

6. AUTHOR(S)
Peter C. Jurs

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Penn State University
152 Davey Laboratory - Chemistry
University Park, PA 16802

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 34412.2-CH

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The specific goal of this project was to develop mathematical models to predict the reduced ion mobility constants, Ko values, for organic compounds directly from their molecular structures. These models are generated by a three-step procedure that involves the representation of the compounds by calculated molecular structure descriptors, selection of the most important descriptors, and the subsequent development of the models using computational neural networks. We have completed and published a high quality model for the prediction of Ko values for monomer ions of 168 compounds using a 6-4-1 (6 input, 4 hidden, and 1 output neuron) computational neural network model. A subset of 93 compounds which exhibited good dimer ion peaks was used to develop a successful 4-2-1 CNN model. A study of phosphorus-containing compounds was also successfully completed. The significance of this work is that it provides fundamental information for ion mobility spectrometry, a sensitive analytical technique used to detect chemical warfare agents in the field.

14. SUBJECT TERMS
Ion Mobility Spectrometry

15. NUMBER IF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Final Progress Report

Prediction of Reduced Ion Mobility Constants from Molecular Structure

ARO Proposal Number: 34414-CH

Period Covered by Report: 1995 – 1998

Title of Proposal: Prediction of Reduced Ion Mobility Constants from Molecular Structure

Grant Number: DAAH04-95-1-0597

Name of Institution: The Pennsylvania State University

Author of Report: Peter C. Jurs

STATEMENT OF THE PROBLEM STUDIED

Research Objectives

The research project involved the development of computational methods to generate mathematical models that link molecular structures of organic compounds and their reduced ion mobility constants, K_o, values. The K_o values are directly observed quantities generated by ion mobility spectrometry (IMS) instruments. IMS is an important analytical chemical method for the determination of extremely low levels of organic compounds.

Approaches to Accomplish Objectives

Relationships between molecular structure and analytical or chemical properties such as K_o can be investigated for large sets of organic compounds using computer-assisted methods. Such quantitative structure-property relationship (QSPR) studies involve three major activities: *representation, feature selection,* and *mapping*. *Representation* involves the calculation of molecular structure descriptors to encode the chemical compounds being studied. Descriptor classes include topological, geometrical, electronic, and hybrid representations of the molecules. Topological descriptors are calculated directly from the connection table representation of the structure, and geometric descriptors are calculated from three-dimensional molecular models. Electronic descriptors come from empirical or molecular orbital calculations. Hybrid descriptors are calculated using several of these representations. *Feature selection* involves selecting the most informative descriptors in the descriptor pool using statistical methods, simulated annealing, or the genetic algorithm. *Mapping* involves analysis of the descriptors using multivariate statistical or computational neural networks to build mathematical models linking the descriptors directly to the property

under investigation, K_o. After their development from a training set, these models then can be used for predicting K_o values for unknown compounds.

The QSPR studies done at Penn State University were done with a specially-developed computer software system designed to provide the capabilities necessary to perform such structure-property relationship studies.

SUMMARY OF THE MOST IMPORTANT RESULTS

A study was done with a set of 168 compounds and their associated K_o values which were provided by Dr. Gary Eiceman of New Mexico State University. The Automated Data Analysis and Pattern recognition Toolkit (ADAPT) software package was the primary software package used in this research.

The 168 compounds were entered and stored and 3-D conformations were generated. A set of 158 numerical descriptors were calculated to encode structural features: 83 topological, 23 geometric, 48 hybrid descriptors. The numerical descriptors were then used to develop linear regression equations and computational neural network models that accurately predicted K_0 for each compound.

Monomer Ion Study: A six-descriptor model was found that accurately calculated K_0 values with a root mean square (rms) error of 0.047 K_0 units. The external prediction set rms error was 0.040 K_0 units, so the model was well validated.

The six descriptors in the linear model were also used to develop a nonlinear computational neural network model. Neural networks take advantage of non-linear relationships that exist between the descriptors and the K_o values. A 6:4:1 (6 input neurons, 4 hidden neurons, 1 output neuron) network architecture was developed, and it had an rms error

of about 0.040 K_0 units for the training set and 0.038 K_0 units for the prediction set, a substantial improvement over the linear model.

This work was published as M. D. Wessel, J. M. Sutter, and P. C. Jurs, "Prediction of Reduced Ion Mobility Constants of Organic Compounds from Molecular Structure," *Analytical Chemistry* **1996**, *68*, 4237-4243.

The ability to predict K_0 values is important since it allows for a better understanding of the structural features that are important to IMS. A computer algorithm could be used in the portable IMS instruments currently in use to aid in the identification of unknowns. Having the ability to predict the K_0 values from structure certainly aids in this development.

Dimer Ion Study: A data set of 93 compounds that exhibited well behaved dimer ion peaks in their mobility spectra was also extracted from the Eiceman data base. The compounds in this data set were also modeled as monomer neutral compounds. The first approximation was that monomer species would adequately encode the features of dimer species. A set of 156 descriptors was calculated. A 4-descriptor linear model was developed for the prediction of K_0 values for dimer ions, and it had rms errors for the training and prediction sets on the order of 0.030 K_0 units. A 4:2:1 neural network improved the errors to about 0.028 K_0 units. This work has been published in M. D. Wessel, "Computer-Assisted Development of Quantitative Structure-Property Relationships and Design of Feature Selection Routines," Ph.D. Thesis, Penn State University, May 1997.

Phosphorus-Containing Compounds: Neither of the two data sets described above included any compounds containing phosphorus. Since the Army is interested in detecting phosphorus-containing compounds, an attempt was then made to model the original

monomer data set supplemented with data for 16 phosphorus-containing compounds. The new, overall data set contained 184 compounds.

One problem that arose concerned the atomic charges on the atoms making up the phosphorus-containing compounds. The ADAPT routine Charge, which was used in the first study of the monomers, is not parameterized for phosphorus-containing compounds. Because of this, several descriptors could not be calculated, as they are dependent on atomic charge information. The semi-empirical molecular orbital package Mopac can calculate charges for phosphorus-containing compounds, so the charge information was extracted from the Mopac output files, and it was used to generate those descriptors that are dependent on charge information. This provided an adequate work-around of the problem of phosphorous compounds and ADAPT.

The 184-compound data set was split randomly into a training set of 166 compounds (including 14 phosphorus-containing compounds) and an external prediction set of 18 compounds (including 2 phosphorus-containing compounds). Once the molecular structure descriptors were calculated, model development commenced. A 7-descriptor linear model was found that calculated K_0 values with an rms error of 0.048 K_0 units. The prediction set rms error was 0.054 K_0 units. The main contribution to the rms error of the training set (0.048 K_0 units) came from compounds that did not contain phosphorus, thus providing more evidence that we could model the K_0 values for compounds that contain phosphorus. The descriptors in this model contained geometric information, which is different from the original 6-descriptor model for monomers which contained no geometric information. These modeling results were encouraging, as it showed that phosphorus-containing com-

pounds could be modeled with only minor improvements to our existing technology. These results were provided to the sponsors in a technical report previously.

Other Ion Mobility Studies: Discussions with Dr. Gary Eiceman of New Mexico State University and Dr. A. Peter Snyder of ERDEC led to a decision to study another, larger set of compounds. The ion mobility spectra of these compounds were to be gathered under differing conditions to assess the effects of the changes on conditions, e.g., moisture content. Due to unavoidable problems with laboratory instrumentation and personnel within the Eiceman laboratory, the availability of these data was very severely delayed. We did received a set of data for 147 compounds in the spring of 1998. The 147 compounds included 10 acids, 14 alcohols, 10 aldehydes, 14 alkanes, 7 alkenes, 14 aromatics, 11 cyclo-alkanes, 10 esters, 10 ketones, 8 mercaptans, 6 nitro-compounds, 16 organic phosphates, 6 phenols, 5 polyaromatics, and 6 sulfides. Don Eldred analyzed these data and it was determined that the ion mobility spectra were not of sufficient quality and reproducibility to allow us to proceed with a QSPR study of these data.

LIST OF ALL PUBLICATIONS AND TECHNICAL REPORTS

M.D. Wessel, J.M. Sutter, P.C. Jurs, "Prediction of Reduced Ion Mobility Constants of Organic Compounds from Molecular Structure," *Analytical Chemistry* 1996, 68, 4237-4243.

P.C. Jurs, "Prediction of Reduced Ion Mobility Constants from Molecular Structure," 3rd International Workshop on Ion Mobility Spectrometry, Galveston, TX 10/14/94

M.D. Wessel and P.C. Jurs, "Prediction of Reduced Ion Mobility Constants from Molecular Structure," 4th International Workshop on Ion Mobility Spectrometry, Cambridge, England, August 6-9, 1995.

M.D. Wessel and P.C. Jurs, "Prediction of Reduced Ion Mobility Constants from Molecular Structure," 5th International Workshop on Ion Mobility Spectrometry, Jackson, WY, August 1996.

M.D. Wessel and P.C. Jurs, "Prediction of Reduced Ion Mobility Constants from Molecular Structure," Scientific Conference on Chemical and Biological Defense Research, Edgewood Area, Aberdeen Proving Ground, MD, Nov. 20, 1996.

LIST OF ALL PARTICIPATING SCIENTIFIC PERSONNEL SHOWING ANY ADVANCED DEGREES EARNED BY THEM WHILE EMPLOYED ON THE PROJECT

Peter C. Jurs, principal investigator

Matthew D. Wessel, graduate student, Ph.D., May 1997

Jonathan M. Sutter, graduate student, Ph.D., Dec. 1997

Donald V. Eldred, graduate student, M.S., May 1999